Title
Evaluation of the Psychometric Properties of the Five Facet of Mindfulness Questionnaire.

Journal
Journal of psychopathology and behavioral assessment, 42(2)

ISSN
0882-2689

Authors
Shallcross, Amanda
Lu, Nathaniel Y
Hays, Ron D

Publication Date
2020-06-01

DOI
10.1007/s10862-019-09776-5

Peer reviewed

Evaluation of the Psychometric Properties of the Five Facet of Mindfulness Questionnaire

Amanda Shallcross, ND, MPH, Nathaniel Y. Lu, MA, Ron D. Hays, PhD

Objective. The Five Facet of Mindfulness Questionnaire (FFMQ) is widely used to assess mindfulness. The present study provides a psychometric evaluation of the FFMQ that includes item response theory (IRT) analyses and evaluation of item characteristic curves.

Method. We administered the FFMQ, the Beck Depression Inventory-II, the Ruminative Response Scale, and the Emotion Regulation Questionnaire to a heterogenous sample of 240 community-based adults. We estimated internal consistency reliability, item-scale correlations, categorical confirmatory factor analysis, and IRT graded response models for the FFMQ. We also estimated correlations among the FFMQ scales and correlations with the other measures included in the study.

Results. Internal consistency reliabilities for the five FFMQ scales were 0.82 or higher. A five-factor categorical model fit the data well. IRT-estimated item characteristic curves indicated that the five response options were monotonically ordered for most of the items. Product-moment correlations between simple-summated scoring and IRT scoring of the scales were 0.97 or higher.

Conclusions. The FFMQ accurately identifies varying levels of trait mindfulness. IRT-derived estimates will inform future adaptations to the

FFMQ (e.g., briefer versions) and the development of future mindfulness instruments.

KEYWORDS: Item Response Theory, Item Characteristic Curves, Mindfulness

Research on the salutary benefits of mindfulness, deliberate and nonjudgmental attentiveness to present-moment experiences (i.e., thoughts, emotions, sensations; Kabat-Zinn 1990), has grown exponentially (Keng et al. 2011). Mounting evidence supports that trait mindfulness and mindfulness training are associated with a range of positive mental and physical health outcomes (Chambers et al. 2016; Shallcross et al. 2015; van Son et al. 2013). One of the most widely used and evaluated measures of mindfulness is the Five Facet of Mindfulness Questionnaire (FFMQ; Baer et al. 2008). The FFMQ is a 39-item instrument that uses five polytomous response options to assess five different aspects of mindfulness: observing, describing, acting with awareness, non-judging, and non-reactivity to inner experience. The majority of the psychometric evaluations of the FFMQ have concentrated on factor analyses and construct validity using classical test theory, with primary exploration around the pattern of intercorrelations between each of the facets, the number of unique and significant factors, factor structure, and the development of shortened versions (e.g., 24-item FFMQ-SF; Bohlmeijer et al. 2011).

For example, several studies have focused on understanding the non-significant correlation between the observing facet and the non-judging facet that has been found among people with prior meditation experience (Baer et al. 2006; Baer et al. 2008; Lilja et al. 2011; Michalak et al. 2016) and whether the observing facet indeed contributes to a higher order mindfulness construct among people without prior meditation experience (Baer et al.

2006; Gu et al. 2016). Additional work has centered around whether the FFMQ is best characterized using a hierarchical or five-factor structure (Gu et al. 2016; Pang and Ruch 2019), which may vary based on exposure to mindfulness interventions and/or meditation experience. Still, the strongest evidence from the general population supports the use of the five-factor model for both the long (Christopher et al. 2012; Veehof 2011) and shortened versions of the FFMQ (Bohlmeijer et al. 2011; Hou 2014). The psychometric properties of the FFMQ-SF and a 15-item short form, the FFMQ-15, have also been reported (Gu et al. 2016).

Given the groundswell of interest in and evidence for the benefits of mindfulness, comprehensive evaluation that extends beyond classical test theory, such as item response theory (IRT) is needed. This is especially necessary for instruments that measure mindfulness given that the definition and operationalization of mindfulness lack clear consensus among experts (Rosch 2007). As a result, the FFMQ, and similar scales, are subject to item miscomprehension and response bias (Van Dam et al. 2010; Van Dam et al. 2009).

Rasch partial credit models have been estimated for the 39-item (long form) and short forms of the FFMQ (Medvedev et al. 2017; Medvedev et al. 2018). Some favor the use of the Rasch model because “it arises from a mathematical formalization of invariance which also turns out to be an operational criterion for fundamental measurement” (Andrich 2004, p. I-15). Others have argued that “it is far better to find models that fit the test data

than to discard data simply to fit the Rasch model" (Hambleton 1994, p. 552). In contrast to the Rasch model, the IRT graded response model (Samejima 1996) estimates item discrimination as well as item difficulty/threshold parameters. The graded response model is more appropriate than the Rasch model because items in a scale frequently differ in how well they represent the underlying construct (item discrimination).

Item characteristic curves (ICCs) describe the relationship between estimated scale scores (latent trait) and responses to items in a scale. ICCs should indicate that each response category has the highest likelihood of being selected within the underlying scale distribution and that the pattern of responses is monotonic—i.e., as an individual's trait level increases, the probability of endorsing an item also increases. ICCs have been used extensively in educational testing applications (Hambleton and Pitoniak 2006; Kaplan and Saccuzzo 1989) and are increasingly recommended for refinement and development of instruments that measure functioning and well-being (Hays et al. 2000; Nguyen et al. 2014; Roberts et al. 2011).

The objective of this study is to evaluate the FFMQ in a heterogeneous sample of 240 individuals. While large samples are ideal for precise measurement of item characteristics and person scores, a "smaller heterogeneous sample is generally preferred to a larger, more homogeneous sample in item parameter estimation" (Hambleton 1994, p. 548). We examine ICCs that describe the relationship between scale scores and response to each FFMQ item. We also evaluate the internal consistency

reliability, item-scale correlations, and construct validity. We hypothesize that factor analyses will support a five-factor model given recent evidence from large-scale studies that support this factor structure across individuals with heterogeneous meditation experience (Christopher et al. 2012; Hou et al. 2014; Pang and Ruch 2019; Veehof 2011). We also hypothesize positive correlations between mindfulness and emotion regulation and negative correlations with depression and rumination.

Method

Participants and Procedures

The analytic sample included 240 people pooled from two separate community-based studies that included participants with elevated levels of self-reported depressive symptoms. Study 1 sample included participants with depressive symptom scores between 4-30 on the Beck Depression Inventory II (BDI-II; Beck et al. 1996), which correspond to minimal to moderate depression. Study 2 sample included participants with scores on the 2-item Patient Health Questionnaire (PHQ-2; Kroenke et al. 2003) ≥ 2, which required participants to have a minimum of mild to moderate levels of depression.

The racial/ethnic composition of the sample was: 51% non-Hispanic White; 28% non-Hispanic Black, 8% Hispanic, and 14% Other. The average age was 36 years (SD = 12.2; range: 19-71 years), 75% were female; and 70% had a college degree.

Respondents completed a battery of self-report questionnaires online using the Qualtrics platform. Participants in Study 1 were recruited through referrals from community mental health centers and local advertisements as a part of a larger study that required in-person office visits. Participants in Study 2 were recruited through ResearchMatch (Harris et al. 2012), a widely-used national recruitment database funded by the National Institutes of Health that includes a pool of prospective research participants who receive notifications for studies for which they may qualify based on study eligibility criteria. All study protocols were approved by the local Institutional Review Board (IRB approval # i15-01042) and procedures were carried out in accordance with the approved guidelines. Informed consent was obtained from all participants.

Measures

Mindfulness. As noted above, the FFMQ assesses five facets of a general tendency to be mindful in daily life: 1) observing, noticing, and attending to sensations, perceptions, thoughts, and feelings; 2) describing/labeling one's experience with words; 3) acting with awareness (i.e., automatic pilot/concentration/non-distraction); 4) non-judging; and 5) non-reactivity to inner experience. Higher scores reflect greater levels of trait mindfulness. Items are rated on a five-point categorical response scale ranging from 1 (never or very rarely true) to 5 (very often or always true). The alpha coefficients of each subscale for the current sample ranged from .82 to .93. In prior studies, individual FFMQ scales had good internal

consistency with Cronbach's alphas ranging from .75 to .93 (Christopher et al. 2012; Baer et al. 2008).

Depression. Depression was assessed with 20-items from the Beck Depression Inventory—II (BDI-II; Beck et al. 1996) measured on a four-point response scale. The suicide item (i.e., BDI 9) was not included due to protection of human subjects' concerns raised by the Institutional Review Board for Study 2. Thus, this item was not included in analyses. Higher scores reflect higher levels of depressive symptoms. The alpha coefficient of the BDI-II demonstrated for the current sample was .87. The BDI-II has an internal consistency of .90 with a retest reliability from .73 to .96 (Beck et al. 1996).

Rumination. Rumination was assessed using the 10-item Ruminative Response Scale (RRS-10; Treynor et al. 2003) measured on a four-point response scale (1=almost never, 2=sometimes, 3=often, 4=almost always). The RRS-10 has been shown to represent two types of rumination: 1) brooding and 2) reflective. Brooding rumination refers to "mood pondering," and reflective rumination involves purposeful and non-emotional reassessment of experienced events, feelings, and behaviors. Higher scores reflect greater tendency to ruminate. The alpha coefficient of the RRS for the current sample was 82. The RRS-10 is highly correlated with the full 22-item version of the scale (r = .90) and has high internal consistency (Cronbach's α = .85) (Erdur-Baker and Bugay 2010; Treynor et al. 2003).

Emotion regulation. Emotion regulation was assessed using six items from the Emotion Regulation Questionnaire (ERQ; Gross and John 2003), measured on a seven-point response scale (1=strongly disagree, 4=neutral, 7=strongly agree). The ERQ assesses an individual's habitual use of cognitive reappraisal (positively reframing distressing emotional experiences). Higher scale scores reflect greater cognitive reappraisal. The alpha coefficient of the cognitive reappraisal scale was .89. The ERQ has good internal consistency and sound convergent and discriminant validity (Ioannidis and Siegling 2015; Gross and John 2003).

Analysis Plan

We first scored the 39 items as 1 = *never or very rarely true*, 2 = *rarely true*, 3 = *sometimes true*, 4 = *often true*, and 5 = *very often or always true*. To be consistent with prior work, we computed item frequencies, and estimated internal consistency reliability (Cronbach's α), and item-scale correlations for the five multi-item scales. We then evaluated the dimensionality of the 39 items using categorical confirmatory factor analysis (CFA). Comparative Fit Index (CFI) values greater than 0.90 indicate acceptable fit between a model and the data. For the root mean square error of approximation (RMSEA), a value of 0.05 or less represents a close fit, 0.08 fair fit, and 0.10 marginal fit (Browne and Cudeck 1993). Next, we fit a graded response model for each of the 5 FFMQ scales to estimate discrimination parameters (a_i) and category threshold parameters (b_i), and item characteristic curves for each item. In the graded response model, each

item is described by a slope parameter (a_i) and between category threshold parameters (b_i), one less than the number of response options. The slope or item discrimination parameter provides an estimate of how well the item differentiates between individuals with varying trait levels. Discrimination parameters are analogous to item-total correlations and typically range from 0 to +2 (Hays et al. 2000). Values of a_i from 0.01 to 0.24 are considered very low, 0.25–0.63 low, 0.65–1.34 moderate, 1.35–1.69 high, and >1.7, very high discrimination (Baker 2001). However, these ranges are context-dependent and can vary by number of response options (e.g., larger slopes would be expected with more response options). Threshold parameters represent the trait level necessary to have a 0.50 probability of responding below versus above the threshold.

Finally, we estimated correlations among the FFMQ facets scored using simple summated scoring and IRT scores estimated from the graded response model. In addition, we computed correlations of the FFMQ facets with several other variables (i.e., depression, rumination, emotion regulation) each of which have been shown to be associated with the FFMQ in prior studies.

The SAS® 9.4 (TSIM3) statistical software (SAS Institute, Inc 2014) was used to estimate correlations, internal consistency reliability, and graded response model parameters. The confirmatory factor analytic model was estimated using Mplus Version 7 (Muthén and Muthén 2014).

Results

Descriptive statistics (means, standard deviations, and ranges) are provided in Table 1 for the Beck Depression Inventory—II, RRS-10, ERQ, and FFMQ scales. Internal consistency reliability estimates for the five facets (scales) were 0.82 or higher: Observing (8 items, alpha = 0.82), Describing (8 items, alpha = 0.93), Acting with Awareness (8 items, alpha = 0.90), Non-judging of inner experience (8 items, alpha = 0.91), and Non-reactivity to inner experience (7 items, alpha = 0.82). Item-scale correlations for the 5 scales are shown in Table 2. Correlations of items with the scale they are in (corrected for item overlap with the scale) were large and exceeded correlations of the items with other scales. A five-factor categorical model fit the data well; comparative fit index (CFI) = 0.926, root mean square error of approximation (RMSEA) = 0.068. All factor loadings and factor correlations were statistically significant. Separate CFA models for each scale also indicated good fit with CFI's ranging from 0.965 to 0.977 and support local independence (all residual correlations were less than an absolute value of 0.20). The lowest CFI was 0.914 for the acting with awareness scale and two of its residual correlations involving Item #34 ("I do jobs or tasks automatically without being aware of what I'm doing"), had local dependency with item #5 ("When I do things, my mind wanders off and I'm easily distracted"; r = -0.232) and item 13 ("I am easily distracted"; r = -0.278).

IRT parameter estimates for the 5 scales are provided in Table 3. Item discrimination parameters (a_i) indicated that the items with the highest

slopes (i.e., items that most clearly differentiate between individuals with varying trait levels of mindfulness) in each scale were: Observing (item 15: "*I pay attention to sensations, such as the wind in my hair or sun on my face.*"), Describing (item 12: "*It's hard for me to find the words to describe what I'm thinking*."), Acting with Awareness (item 38: "*I find myself doing things without paying attention*."), Non-Judging (item 25: "*I tell myself that I shouldn't be thinking the way I'm thinking*."), and Non-reactivity (item 29: "*When I have distressing thoughts or images I am able just to notice them without reacting*."). Importantly, each of these items not only reflect the largest a_i within each scale but they also have latent trait scores that were relatively evenly distributed across the trait range (see Table 3). These relatively large discrimination parameters indicate that these items (#15 (Observe); #12 (Describe); #38 (Act with Awareness); #25 (Non-judge); #29 (Non-react)) best differentiate individuals at low through high trait levels of mindfulness within each scale.

Threshold parameters (b_i) ranged from -4.43 to 2.74 (Observing), -3.06 to 1.42 (Describing), -2.44 to 2.49 (Act with Awareness), -2.05 to 1.73 (Non-judging), and -2.92 to 5.46 (Non-reactivity). These results indicate that the items capture a wide range of the underlying mindfulness distribution.

Item characteristic curves for the 39 FFMQ items are given in Figure 1. The curves show the probability of picking each response choice on the y-axis as a function of underlying mindfulness on the x-axis. The five response categories are appropriately monotonically ordered and working as desired

for most of the items because each category is most likely to be selected for some level of underlying level of mindfulness. One exception is item #11 where the second response option (*rarely true*) is never most likely to be selected across the observing scale continuum. Because the deviations from assumptions of the IRT model were minimal and inconsequential, we computed IRT scores using all available response options (i.e., did not collapse options together).

Product-moment correlations within the five FFMQ scales scored using simple summated scoring and IRT scoring, and correlations of these scores with depression, rumination, and emotion regulation are presented in Table 4. Correlations within the five scales using simple summated scoring ranged from a non-significant correlation between Observing and Non-judging to r = 0.47 between Acting with Awareness and Non-judging. A similar pattern was observed for the within scale correlations for IRT scores, whereby we observed a non-significant correlation between Observing and Non-judging and a high correlation between Acting with Awareness and Non-judging (r = 0.48). Correlations between corresponding scales (e.g., simple summated "Observe" with IRT estimated "Observe") ranged from 0.97-0.99. Correlations between each of the five scales and psychological variables (depression, rumination, emotion regulation) were similar for both simple summated scores and IRT scores and were in directions consistent with the literature (e.g., positive correlations between mindfulness and emotion regulation and negative correlations with depression and rumination). Most

of the correlations were statistically significant except the simple summated scored Describe scale relationship with rumination and the Observe scale for both simple summated and IRT scores was not related to rumination.

Discussion

Cronbach's alphas for the five scales and item-scale correlations support internal consistency reliability and item convergence within scales of the FFMQ in a racially/ethnically diverse sample. Confirmatory factor analyses indicated that the five-factor model provided adequate fit to the data.

In this study, we found that item characteristic curves (Figure 1) from IRT graded response models demonstrate that most items were monotonically ordered. One exception to adequate item performance was Item #11 ("*I notice how foods and drinks affect my thoughts, bodily sensations, and emotions.*").The second response option (*rarely true*) was never most likely selected. The modest item discrimination parameter (a_i = .83) for item #11 adds additional verification that this item contributes limited information in terms of its ability to discriminate between individuals with varying levels of trait mindfulness. This and other indications of variation in item discrimination support the use of an IRT model that allows item discriminations to vary, rather than to impose the Rasch model assumption of equal item discriminations.

Overall item parameter estimates indicate that the following items in each scale best differentiate individuals at low through high levels of

mindfulness: (#15 (Observe); #12 (Describe); #38 (Act with Awareness); #25 (Non-judge); #29 (Non-react)). One of the reasons it is helpful to understand which items yield the most information is because the FFMQ is a long measure at 39 items, thus limiting its use in studies where participant burden is a concern. A shorter form of the FFMQ that includes these items and excludes item #11 may represent the most efficient measure of mindfulness.

The FFMQ-SF includes each of the maximally performing items from the 39-item FFMQ and does not include the poorly performing item ("*I notice how foods and drinks affect my thoughts, bodily sensations, and emotions*."). The 15-item version includes each of the ideally performing items, except for #12 (Describe): *"It's hard for me to find the words to describe what I'm thinking."* The FFMQ-15 also includes the worst performing item #11 ("*I notice how foods and drinks affect my thoughts, bodily sensations, and emotions*."). A shortened form of the original, English version of the FFMQ has not yet been evaluated in a non-clinical population. Thus, the item performance information from these IRT analyses may help inform the validation of a brief version of FFMQ that is apt for use in the general U.S. population.

The correlations among the FFMQ scales were significant (except the Observe scale) and ranged from 0.27 to 0.48, suggesting that they represent related but distinct constructs. The Observe scale was not significantly related to the Non-judge scale. Other studies have similarly found non-

significant relationships between Observing and Non-judging (Lilja et al. 2016). It has been suggested that this finding may reflect varying levels of experience with meditation, whereby meditators may have acquired the skills to remain non-judgmental while paying attention to present moment experiences (observing), but those with more limited meditation experience tend to observe in a more judgmental way (Baer 2016; Baer et al. 2006; Baer et al. 2008; Lilja et al. 2016). Other explanations include the potential of constrained scaling of the FFMQ (i.e., ceiling effects, skewness, and non-normal distributions that have been found among both mediating and non-nonmeditating samples) (Christopher et al. 2012; Gu et al. 2016; Pang and Ruch 2019).

The high correlations (0.97 to 0.99) between the simple summated scores and IRT scores were not surprising given that the IRT scores are estimated by optimally weighted items. Correlations between the FFMQ scales and psychological variables were similar for simple summated scores and in expected directions. The positive correlations between mindfulness scales and emotion regulation and negative correlations with depression and rumination in this study are consistent with prior research. The acting with awareness and non-judging of inner experience scales were the most strongly associated with depression, non-judging of inner experience was most strongly correlated with rumination, and non-reactivity to inner experience was most highly related to emotional regulation. The pattern of correlations provides further support for the construct validity of the FFMQ.

This study illustrates the utility of IRT in evaluation and refinement of measures of psychological functioning and health outcomes. Specifically, this study depicted item parameter estimates and ICCs that have not been previously reported for a widely used measure of trait mindfulness. Finally, data were collected from a heterogeneous sample and included assessments of a range of clinically relevant outcomes that have been associated with the FFMQ in prior studies.

Several limitations are worth noting. First, this study did not include any other measure of mindfulness besides the FFMQ. Thus, correlations with other mindfulness scales could not be reported. Second, the sample size for this study was not large (Cappelleri et al. 2014). Third, it is possible that given inclusion criteria for this study (e.g., elevated levels of depressive symptoms) that results may not generalize to a healthy sample. Further, the sample was primarily comprised of females (75%) and were highly educated, which may further limit generalizability. Finally, despite the benefit of including a racially heterogeneous sample (approximately 50% Black participants), differences in cultural relevance of mindfulness (Woods-Giscombe 2014) may have resulted in variance in interpretability of questions between White and Black participants. Future studies with larger samples are needed to replicate results and to compare item parameter estimates and ICCs by race, education, and gender.

In summary, this study is the first to report item discrimination and item difficulty/threshold parameters for the 39-item FFMQ. It provides

further evidence supporting the psychometric properties of this widely used measure. Item parameter estimates and item characteristic curves demonstrate adequate performance of the majority of the 39 items and support that the FFMQ is able to accurately identify individuals with varying levels of trait mindfulness. The graded response model revealed several items that differed in how well they represent the underlying construct of mindfulness. This information has not been previously reported the FFMQ. Understanding which items in the FFMQ that most accurately identify trait levels of mindfulness can inform future adaptations the FFMQ (e.g., shortened versions) and the development of other trait mindfulness measures to most precisely measure this construct.

Compliance with Ethical Standards

The authors declare that they have no conflict of interest in association with this manuscript. All study protocols were approved by the local Institutional Review Board (IRB approval # i15-01042) and procedures were carried out in accordance with the approved guidelines. Informed consent was obtained from all participants.

References

Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care*, 42(1), I7–I16.

Baer, R. (2016). Assessment of mindfulness and closely related constructs: Introduction to the special issue. *Psychological Assessment,* 28(7), 787. http://dx.doi.org/10.1037/pas0000309

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, 13(1), 27-45. https://doi.org/10.1177/1073191105283504

Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., et al. (2008). Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. *Assessment*, 15(3), 329–342. https://doi.org/10.1177/1073191107313003

Baker, F. B. (2001). *The basics of item response theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. Retrieved from https://eric.ed.gov/?id=ED458219

Beck, A. T., Brown, G. K., & Steer, R. A. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Bohlmeijer, E., Ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the Five Facet Mindfulness Questionnaire in depressed adults and development of a short form. *Assessment*, 18(3), 308–320. https://doi.org/10.1177/1073191111408231

Browne, M. W., & Cudeck, R. (1993). *Alternative ways of assessing model fit*. (K. A. Bollen & J. S. Long, Eds.), *Testing structural equation models* (Vol. 154). Newbury Park, CA: Sage. https://doi.org/10.1177/0049124192021002005

Cappelleri, J. C., Jason Lundy, J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, 36(5), 648–662. https://doi.org/10.1016/j.clinthera.2014.04.006

Chambers, S. K., Foley, E., Clutton, S., McDowall, R., Occhipinti, S., Berry, M., et al. (2016). The role of mindfulness in distress and quality of life for men with advanced prostate cancer. *Quality of Life Research*, 25(12), 3027–3035. https://doi.org/10.1007/s11136-016-1341-3

Christopher, M. S., Neuser, N. J., Michael, P. G., & Baitmangalkar, A. (2012). Exploring the psychometric properties of the Five Facet Mindfulness Questionnaire. *Mindfulness,* 3(2), 124-131. https://doi.org/10.1007/s12671-011-0086-x

Erdur-Baker, O., & Bugay, A. (2010). The short verison of Ruminative Response Scale: Reliability, validity and its relation to psychological symptoms. *Procedia-Social and Behavioral Sciences,* 5, 2178-2181. https://doi.org/10.1016/j.sbspro.2010.07.433

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-

being. *Journal of Personality and Social Psychology*, 85(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Gu, J., Strauss, C., Crane, C., Barnhofer, T., Karl, A., Cavanagh, K., & Kuyken, W. (2016). Examining the factor structure of the 39-item and 15-item versions of the Five Facet Mindfulness Questionnaire before and after mindfulness-based cognitive therapy for people with recurrent depression. *Psychological Assessment*, 28(7), 791–802. https://doi.org/10.1037/pas0000263

Hambleton, R. K. (1994). Item response theory: a broad psychometric framework for measurement advances. *Psicothema*, 6(3), 535–556.

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433–470). Westport, CT: Praeger.

Harris, P. A., Scott, K. W., Lebo, L., Hassan, N., Lighter, C., & Pulley, J. (2012). ResearchMatch: A national registry to recruit volunteers for clinical research. *Academic Medicine: Journal of the Association of American Medical Colleges*, 87(1), 66. https://doi.org/10.1097/ACM.0b013e31823ab7d2

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care*, 38(9), 28–42. https://doi.org/10.1016/j.bbi.2008.05.010

Hou, J., Wong, S. Y. S., Lo, H. H. M., Mak, W. W. S., & Ma, H. S. W. (2014). Validation of a Chinese version of the Five Facet Mindfulness

Questionnaire in Hong Kong and development of a short form. *Assessment,* 21(3), 363-371. https://doi.org/10.1177/1073191113485121

Ioannidis, C. A., & Siegling, A. B. (2015). Criterion and incremental validity of the Emotion Regulation Questionnaire. *Frontiers in Psychology*, 6(247). https://doi.org/10.3389fpsyg.2015.00247

Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain and illness.* New York, NY: Dell.

Kaplan, R. M., & Saccuzzo, D. P. (1989). *Psychological testing: Principles, applications, and applications* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Keng, S. L., Smoski, M. J., & Robins, C. J. (2011). Effects of mindfulness on psychological health: A review of empirical studies. *Clinical Psychology Review*, 31(6), 1041-1056. https://doi.org/10.1016/j.cpr.2011.04.006

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care,* 41(11), 1284-1292. https://doi.org/10.1097/01.MLR.0000093487.78664.3C

Lilja, J. L., Frodi-Lundgren, A., Hanse, J. J., Josefsson, T., Lundh, L. G., Sköld, C., et al. (2011). Five facets mindfulness questionnaire—reliability and factor structure: A Swedish version. *Cognitive Behaviour Therapy,* 40(4), 291-303. https://doi.org/10.1080/16506073.2011.580367

Lilja, J. L., Zelleroth, C., Axberg, U., & Norlander, T. (2016). Mindfulness-

based cognitive therapy is effective as relapse prevention for patients with recurrent depression in Scandinavian primary health care. *Scandinavian Journal of Psychology*, 57(5), 464–472. https://doi.org/10.1111/sjop.12302

Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017). Improving the precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mindfulness*, 8(4), 995-1008. https://doi.org/10.1007/s12671-016-0676-8

Medvedev, O. N., Titkova, E. A., Siegert, R. J., Hwang, Y. S., & Krägeloh, C. U. (2018). Evaluating short versions of the Five Facet Mindfulness Questionnaire using Rasch analysis. *Mindfulness*, 9(5), 1411-1422. https://doi.org/10.1007/s12671-017-0881-0

Michalak, J., Zarbock, G., Drews, M., Otto, D., Mertens, D., Stroehle, G., et al. (2016). Assessment of mindfulness with the German version of the Five Facet Mindfulness Questionnaires (FFMQ-D). *Zeitschrift für Gesundheitspsychologie*, 24(1), 1-12. https://doi.org/10.1026/0943-8149/a000149

Muthén, L. K., & Muthén, B. O. (2014). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén. Retrieved from http://www.statmodel2.com/download/usersguide/MplusUserGuideVer_7.pdf

Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The*

Patient, 7(1), 23–35. https://doi.org/10.1007/s40271-013-0041-0

Pang, D., & Ruch, W. (2019). Scrutinizing the components of mindfulness: Insights from current, past, and non-meditators. *Mindfulness*, 10(3), 492-505. https://doi.org/10.1007/s12671-018-0990-4

Roberts, M. E., Barthel, F. M. S., Lopez, C., Tchanturia, K., & Treasure, J. L. (2011). Development and validation of the Detail and Flexibility Questionnaire (DFlex) in eating disorders. *Eating Behaviors*, 12(3), 168–174. https://doi.org/10.1016/j.eatbeh.2011.04.001

Rosch, E. (2007). More than mindfulness: When you have a tiger by the tail, let it eat you. *Psychological Inquiry,* 18(4), 258-264. https://doi.org/10.1080/10478400701598371

Samejima, F. (1996). Graded response model. In *Handbook of modern item response theory* (pp. 85–100). New York, NY: Springer.

SAS Institute Inc. (2014). *Base SAS® 9.4 procedures guide: Statistical procedures*. Cary, NC: SAS Institute Inc.

Shallcross, A. J., Gross, J. J., Visvanathan, P. D., Kumar, N., Palfrey, A., Ford, et al. (2015). Relapse prevention in major depressive disorder: Mindfulness-based cognitive therapy versus an active control condition. *Journal of Consulting and Clinical Psychology*, 83(5), 964–975. https://doi.org/10.1037/ccp0000050

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*, 27(3), 247–259. https://doi.org/10.1023/A:1023910315561

Van Dam, N. T., Earleywine, M., & Borders, A. (2010). Measuring mindfulness? An item response theory analysis of the Mindful Attention Awareness Scale. *Personality and Individual Differences*, 49(7), 805–810. https://doi.org/10.1016/J.PAID.2010.07.020

Van Dam, N. T., Earleywine, M., & Danoff-Burg, S. (2009). Differential item function across meditators and non-meditators on the Five Facet Mindfulness Questionnaire. *Personality and Individual Differences*, 47(5), 516–521. https://doi.org/10.1016/J.PAID.2009.05.005

van Son, J., Nyklícek, I., Pop, V. J., Blonk, M. C., Erdtsieck, R. J., Spooren, et al. (2013). The effects of a mindfulness-based intervention on emotional distress, quality of life, and HbA(1c) in outpatients with diabetes (DiaMind): A randomized controlled trial. *Diabetes Care*, 36(4), 823–830. https://doi.org/10.2337/dc12-1477

Veehof, M. M., Peter, M., Taal, E., Westerhof, G. J., & Bohlmeijer, E. T. (2011). Psychometric properties of the Dutch Five Facet Mindfulness Questionnaire (FFMQ) in patients with fibromyalgia. *Clinical Rheumatology,* 30(8), 1045-1054. https://doi.org/10.1007/s10067-011-1690-9

Woods-Giscombé, C. L., & Gaylord, S. A. (2014). The cultural relevance of mindfulness meditation as a health intervention for African Americans: Implications for reducing stress-related health disparities. *Journal of Holistic Nursing,* 32(3), 147-160. https://doi.org/10.1177/0898010113519010